Top cross section measurement in the lepton+jets channel with b-tagging at DØ

Alexander Khanov

University of Rochester

for the DØ collaboration

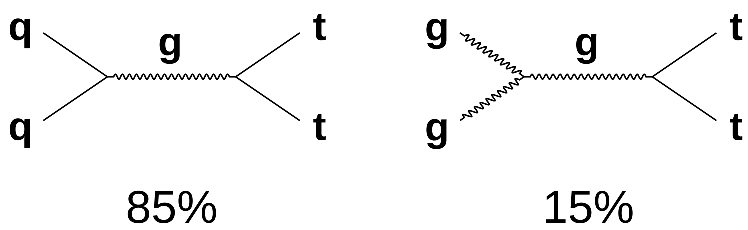
W&C Seminar, September 10, 2004

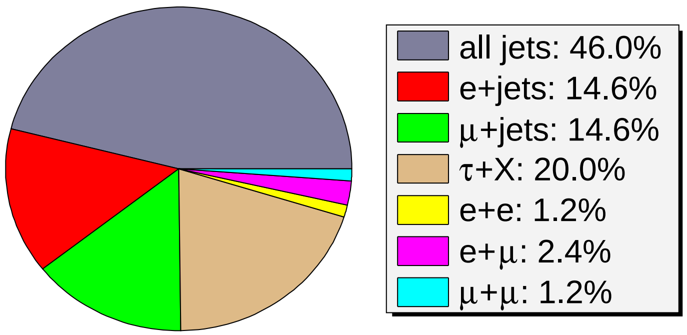
Motivation

- the knowledge on top quark properties from Run I data has been severely affected by the small statistics available. This should not be an issue in Run II with the large anticipated data samples
- the measurement of the $t\bar{t}$ cross section is a good test of perturbative QCD
- New Physics may manifest itself in top production (e.g. $t\bar{t}$ resonance) or decay (e.g. $t \rightarrow H^+ b$)
- studies of top production are important in the LHC perspective where $t\bar{t}$ is a dominant background to many searches for new physics

$t\bar{t}$ production in lepton+jets mode

- at the Tevatron, top quarks are mostly produced in pairs ($t\bar{t}$)



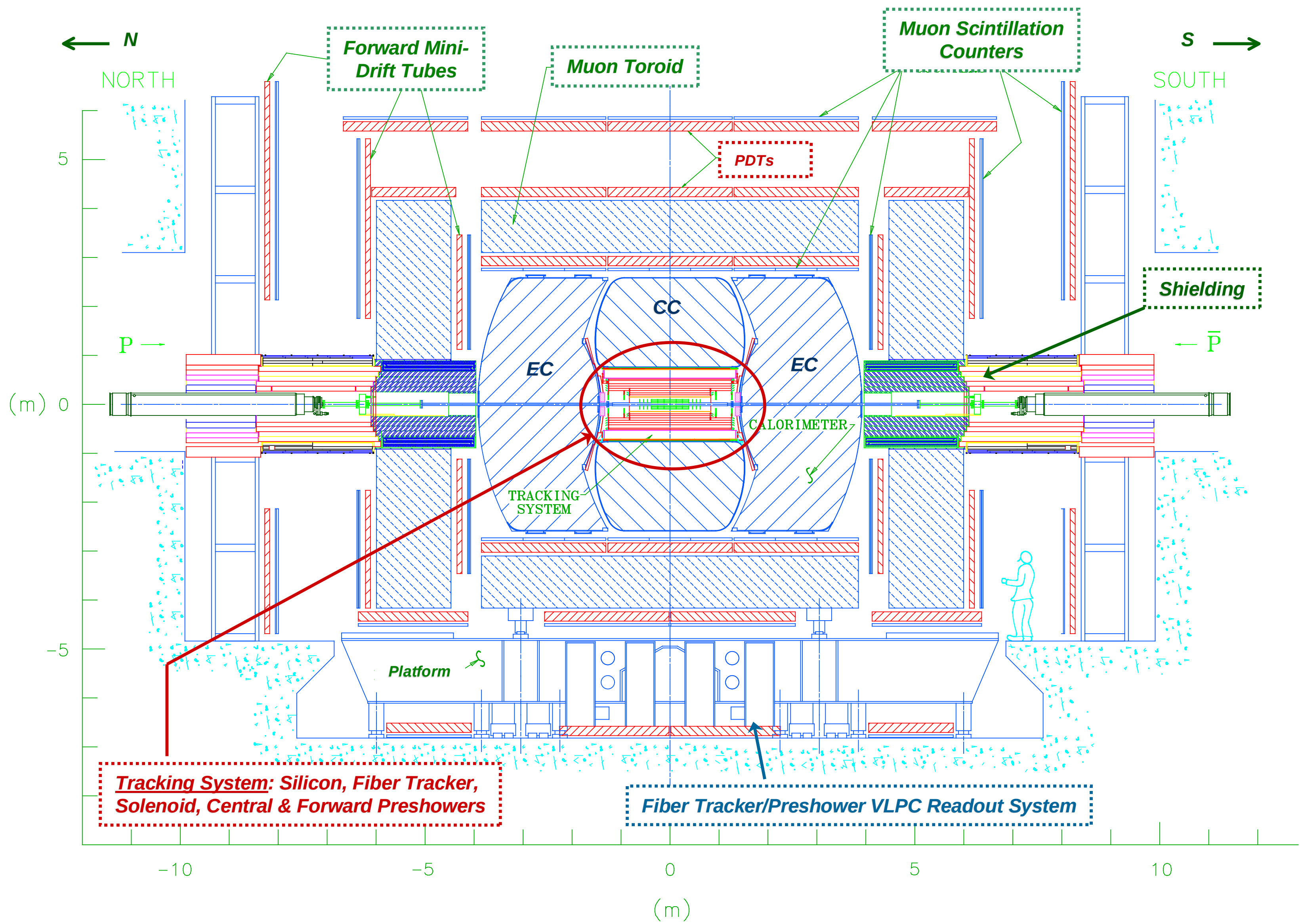


- in the Standard Model, top almost always decays to Wb
- as W may decay hadronically or leptonically, there are dilepton, lepton+jets, and all jets $t\bar{t}$ channels

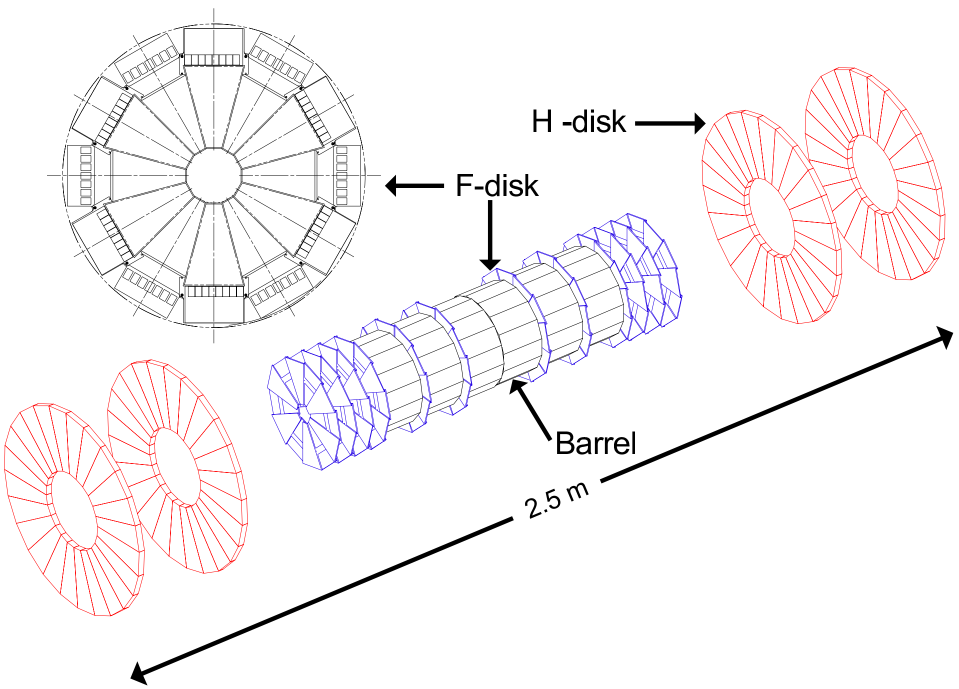
- look at the lepton+jets channel:
  - large statistics compared to dileptons
  - clear signature compared to all jets

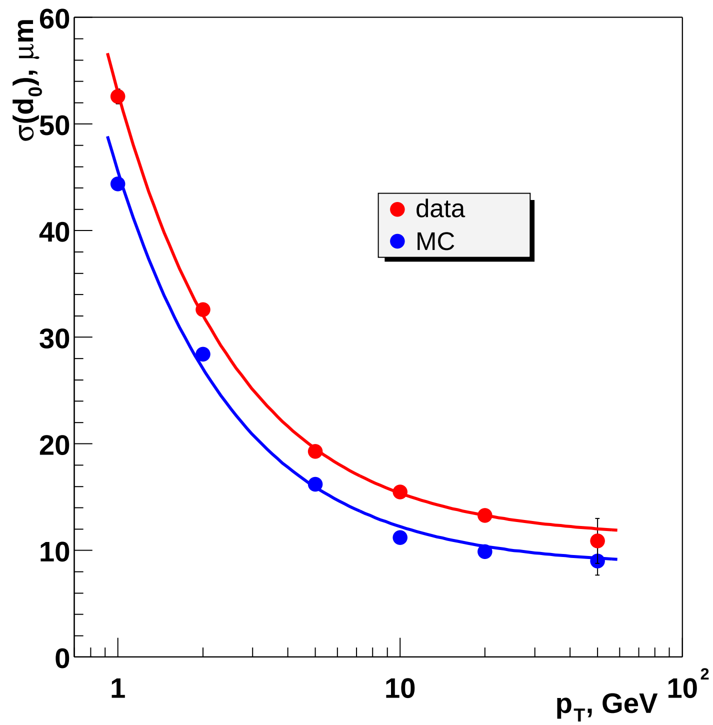
- the purity of the lepton+jets channel is not that high as for dileptons, need a method to increase the fraction of the signal
- one approach used since Run I is the topological selection
- in the present analysis use b-tagging (b-jet identification)

DØ detector in Run II



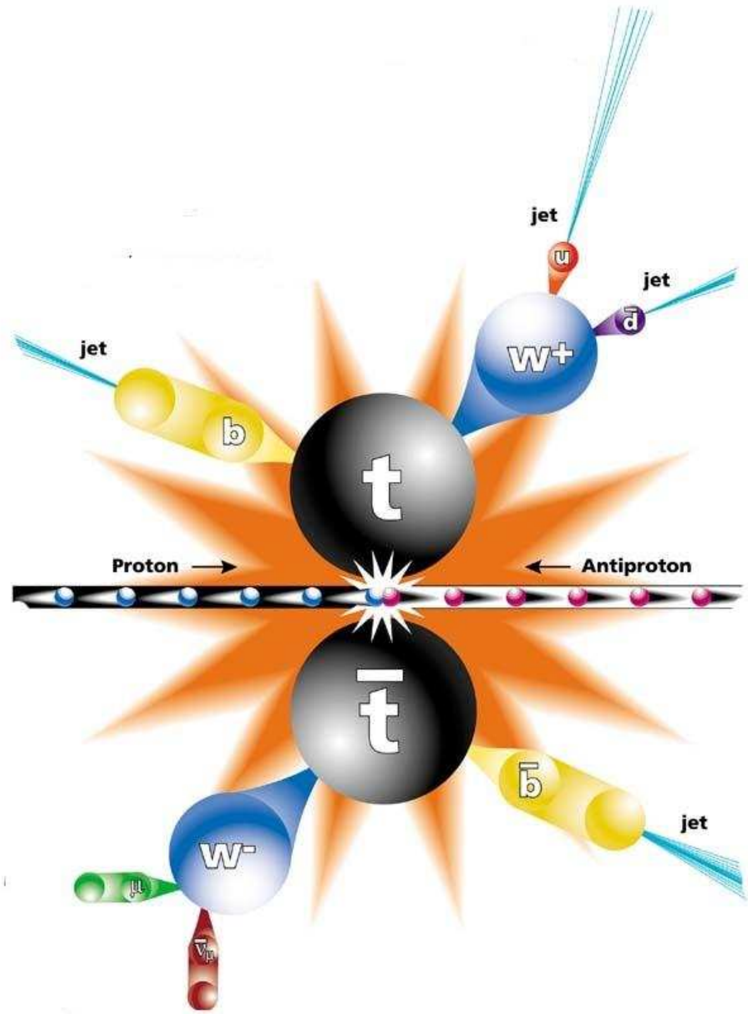
Silicon Microstrip Tracker





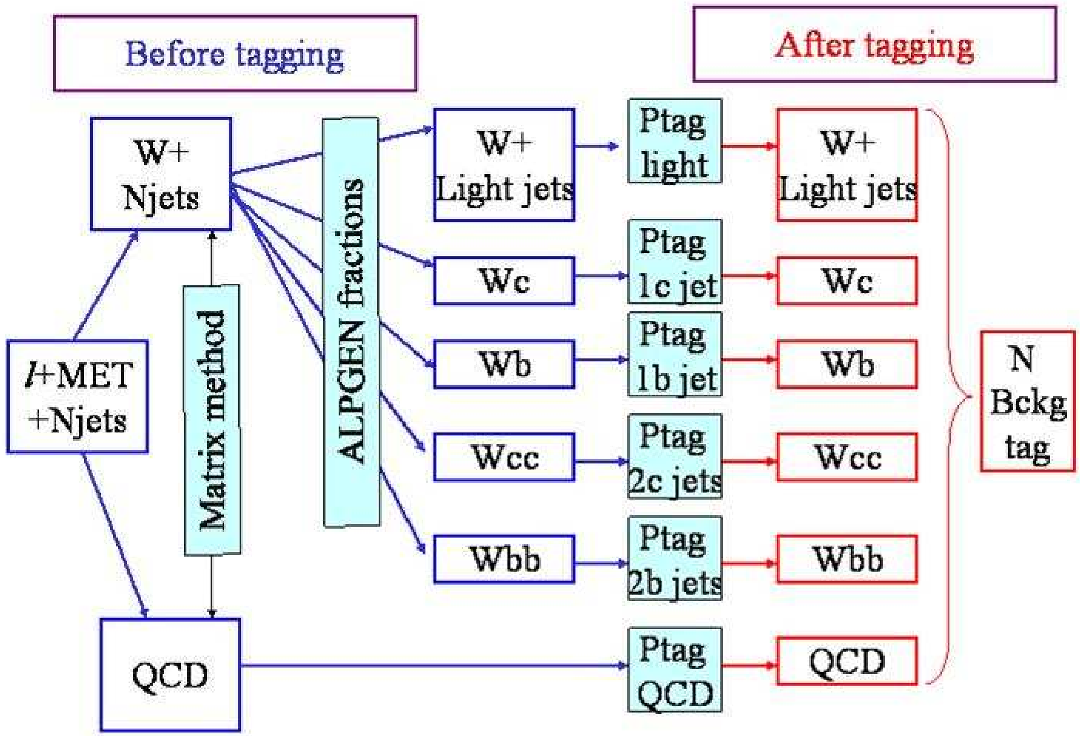
impact parameter resolution in data: $\sigma(d_0) = 11 + 42\ \mathrm{GeV}/p_T\ \mu\mathrm{m}$

Event selection



- The signature:
  - a lepton+jet trigger
  - a $p_T >$20 GeV isolated lepton within $|\eta| <$1.1 (e) or $|\eta| <$2 (μ)
  - high missing E_T (20 GeV for e+jets, 17 GeV for μ+jets)
  - at least three jets with $p_T >$15 GeV, $|\eta| <$2.5
- Additional requirements:
  - high quality primary vertex ($N_{tr} \geq$3, $|z| <$60 cm)
  - triangular cut in $\Delta\varphi(l, \not{E}_T)$ vs $\not{E}_T$
  - second high p_T lepton isolation veto

Background estimation



$$
\begin{aligned}
N &= N_{t\bar{t}} + N_W + N_{QCD} + N_S \\
N^{tag} &= N^{tag}_{t\bar{t}} + N^{tag}_W + N^{tag}_{QCD} + N^{tag}_S
\end{aligned}
$$

- subtract small backgrounds (single top, VV, $Z \to \tau^+\tau^-$) using known cross sections
- calculate QCD (non-W) contribution with matrix method
- separate W from $t\bar{t}$ using difference in their event tagging probabilities
- $t\bar{t}$ signal is observed as an excess in tagged events with ≥ 3 jets over background prediction

QCD background estimation: matrix method

- define a "loose" sample by relaxing the lepton quality cuts
- derive the QCD fraction from the measured probabilities for a "true" lepton ($\varepsilon_{W+t\bar{t}}$) and a "fake" lepton (ε_{QCD}) to go from the loose to the tight sample

$$\begin{aligned} N^{loose} &= N^{loose}_{QCD} + N^{loose}_{W+t\bar{t}} \\ N^{tight} &= \varepsilon_{QCD} N^{loose}_{QCD} + \varepsilon_{W+t\bar{t}} N^{loose}_{W+t\bar{t}} \end{aligned}$$

- method I: apply matrix method to the untagged sample, then

$$N^{tag}_{QCD} = P_{QCD} N_{QCD}$$

  P_{QCD}: probability to tag a QCD event, measured in data

- method II: apply matrix method directly to the tagged sample

- e+jets: QCD dominated by jets faking electrons
  - both methods are shown to give compatible results, use method I as having superior statistical precision

- μ+jets: QCD dominated by muons from semileptonic heavy flavor decays
  - have to use method II in absence of a reliable estimation for P_{QCD} in this case (QCD HF composition is different for low and high $\not{E}_T$)
  - caveat: smaller number of events on tagged sample leads to relatively large statistical error on N^{tag}_{QCD}

Event tagging probabilities

- this is the core of the analysis: separation of $t\bar{t}$ from the W+jets background is based on difference between their event tagging probabilities
- use kinematics of events from Monte Carlo
- obtain tagging probabilities on data, parameterize, and apply to jets in Monte Carlo in form of tag rate functions ε_J according to relevant jet flavors (J = b, c, light):

probability to have no tags: $P_n^{NT} = \left\langle \prod_{k=1}^{n} (1-\varepsilon_{J_k}(E_{Tk},\eta_k)) \right\rangle$

probability to have one tag: $P_n^{ST} = \left\langle \sum_{m=1}^{n} \varepsilon_{J_m}(E_{Tm},\eta_m) \prod_{k\neq m} (1-\varepsilon_{J_k}(E_{Tk},\eta_k)) \right\rangle$

probability to have $\geq$ 2 tags: $P_n^{DT} = 1 - P_n^{NT} - P_n^{ST}$

- probabilities averaged over MC samples implementing relevant corrections that might affect event topology (e.g. trigger efficiency)

Calculation of tagging effi ciencies

- general approach: begin with quantities measured in data, convert them to ε_J using scale factors (SF) derived on Monte Carlo

$$b\text{-tagging efficiency:} \quad \varepsilon_b = \varepsilon_b^{tagg} \varepsilon_{b\to\mu}^{data} SF_{b\to\mu}^{b}$$

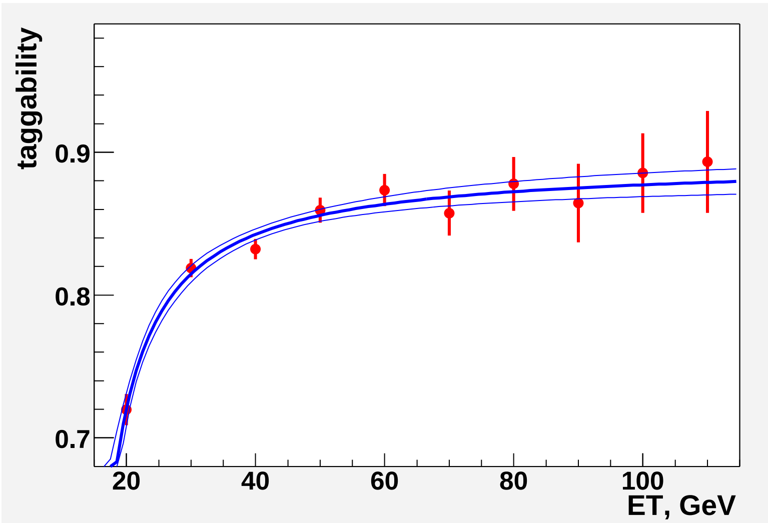
$$c\text{-tagging efficiency:} \quad \varepsilon_c = \varepsilon_c^{tagg} \varepsilon_{b\to\mu}^{data} SF_{b\to\mu}^{c}$$

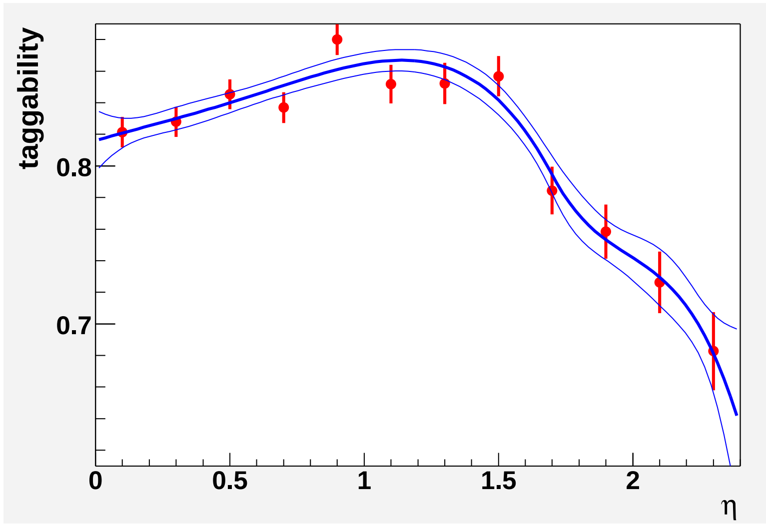
$$\text{mis-tagging rate:} \quad \varepsilon_l = \varepsilon_l^{tagg} \varepsilon_{neg}^{data} SF_l$$

- ε_J^{tagg}: taggability (probability for a jet to be taggable) measured in data and corrected for heavy flavor jets using factors derived on Monte Carlo
- $\varepsilon_{b\to\mu}^{data}$: b-tagging efficiency measured in data for jets with a muon inside
- ε_{neg}^{data}: negative tagging rate measured on data

Taggability

- only a jet that satisfies certain requirements on the number and minimum p_T of tracks associated with it can be tagged. These jets are called “taggable”
- taggability: the probability for a jet to be taggable
- the idea is to largely decouple the tagging efficiency from issues related to tracking inefficiencies and/or calorimeter noise problems
- taggability depends on event sample and is not fully modeled by Monte Carlo

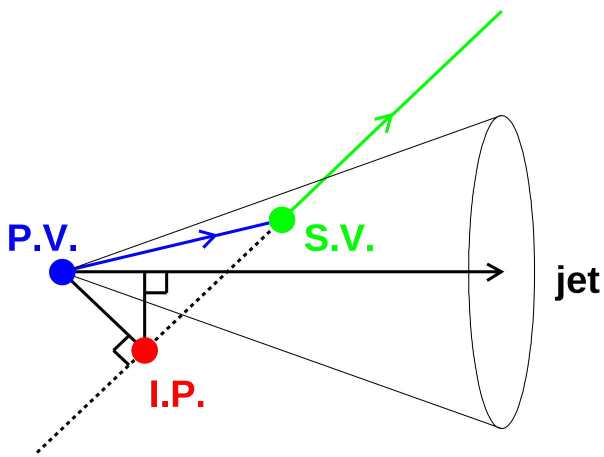




taggability derived on preselected e+jets sample

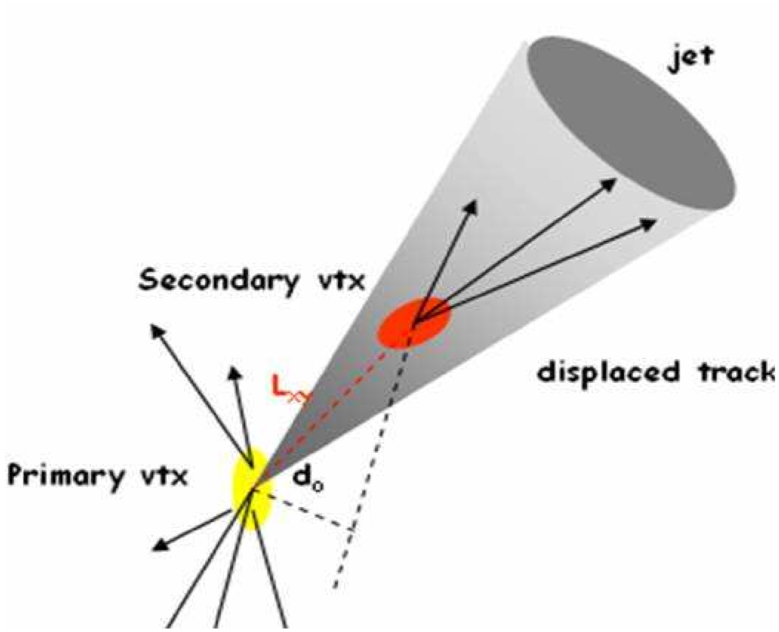
Lifetime b-tagging methods

Counting Signed Impact Parameter (CSIP)



- count the number of tracks with large positive DCA significance S
- jet is positively tagged if it has
  - at least two tracks with $S > 3$, or
  - at least three tracks with $S > 2$

Secondary Vertex Tagger (SVT)



- explicitly reconstruct 3d vertices out of tracks in track-jets using build-up algorithm
- jet is tagged as a b-jet if the signed decay length significance L_{xy} >7

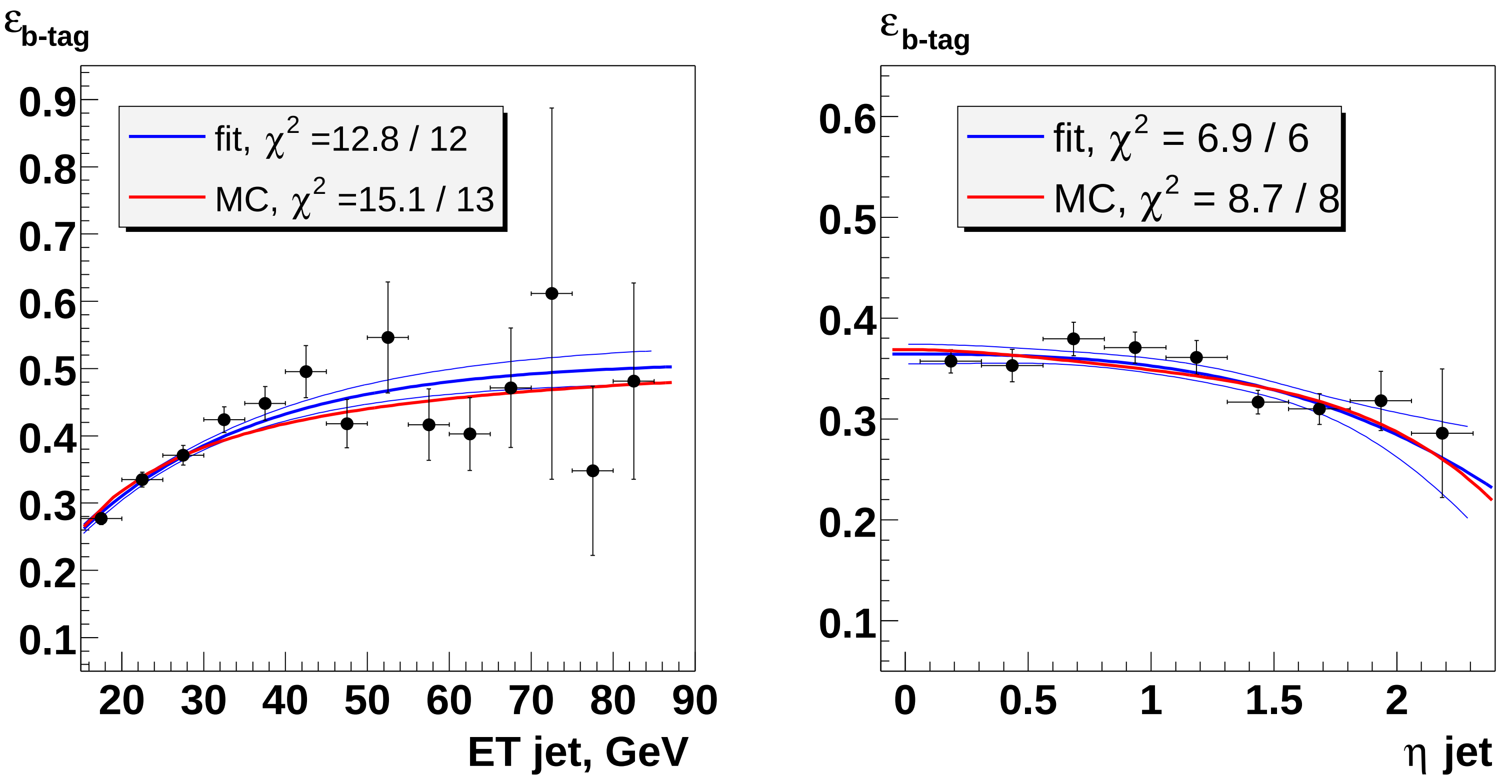
b-tagging effi ciency estimation: System 8

- measure b-tagging efficiency in data for jets with a muon inside
- have two samples with different heavy flavor fractions (increased by tagging the away jet)
- tag jets with two independent tagging algorithms: lifetime tag (LT = CSIP, SVT) and soft lepton tag (SLT = a muon with p_T^{rel} >0.7 GeV inside a jet)

$$
\begin{aligned}
n &= n_b + n_l \\
p &= p_b + p_l \\
n^{LT} &= n_b \varepsilon_{btag}^{LT} + n_l \varepsilon_{non-btag}^{LT} \\
p^{LT} &= p_b \varepsilon_{btag}^{LT} + p_l \varepsilon_{non-btag}^{LT} \\
n^{SLT} &= n_b \varepsilon_{btag}^{SLT} + n_l \varepsilon_{non-btag}^{SLT} \\
p^{SLT} &= p_b \varepsilon_{btag}^{SLT} + p_l \varepsilon_{non-btag}^{SLT} \\
n^{both} &= n_b \varepsilon_{btag}^{LT} \varepsilon_{btag}^{SLT} + n_l \varepsilon_{non-btag}^{LT} \varepsilon_{non-btag}^{SLT} \\
p^{both} &= p_b \varepsilon_{btag}^{LT} \varepsilon_{btag}^{SLT} + p_l \varepsilon_{non-btag}^{LT} \varepsilon_{non-btag}^{SLT}
\end{aligned}
$$

- dominant sources of systematics:
  - assumption on decorrelation between LT and SLT
  - assumption on independence of the b-tagging probability on whether or not the away jet is tagged

b-tagging effi ciency in data



b-tagging effi ciency in data (CSIP), similar effi ciency measured for SVT

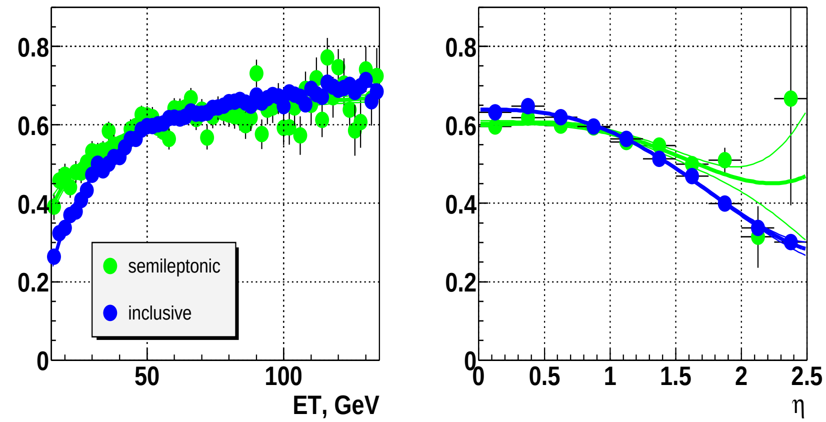
- the shape of the *b*-tagging efficiency is predicted by Monte Carlo
- the absolute value of the Monte Carlo prediction on the plot is normalized to data
- use the error band of the data fit to estimate the error on the *b*-tagging efficiency

b,c-tagging efficiency in Monte Carlo and scale factors

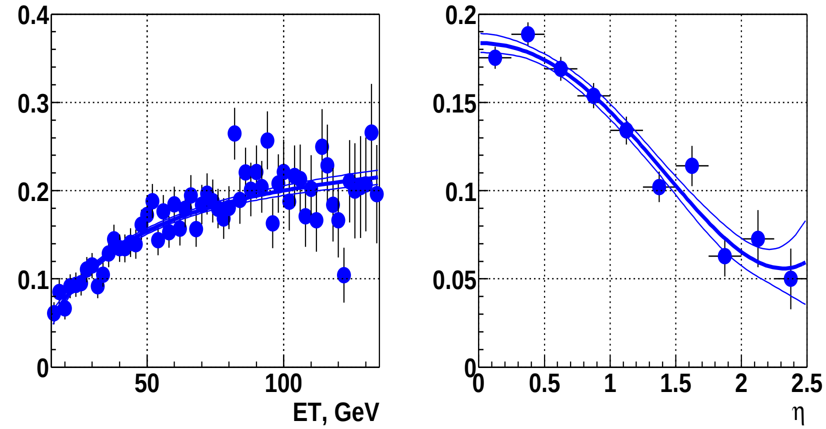
- what is measured in data is semileptonic b-tagging efficiency
- need inclusive b-tagging efficiency and inclusive c-tagging efficiency
- derive relevant scale factors on Monte Carlo

  $SF^b_{b\to\mu} = \varepsilon_b/\varepsilon_{b\to\mu}$: crucial, but close to 1 except for low jet E_T

  $SF^c_{b\to\mu} = \varepsilon_c/\varepsilon_{b\to\mu}$: significantly different from 1, but does not affect much the result



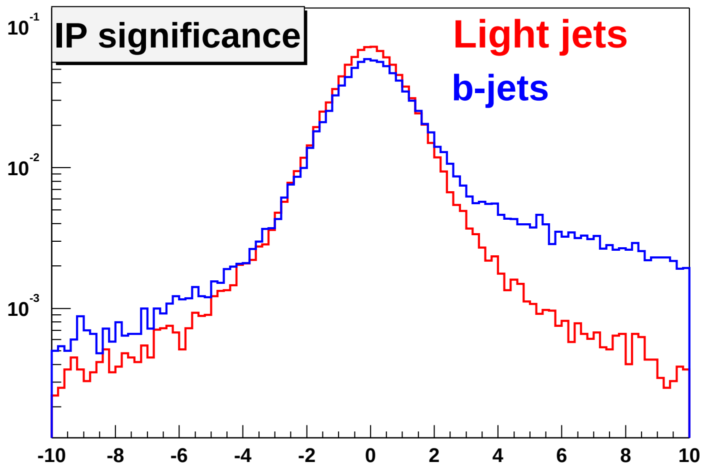
b-tagging efficiency in MC (CSIP)

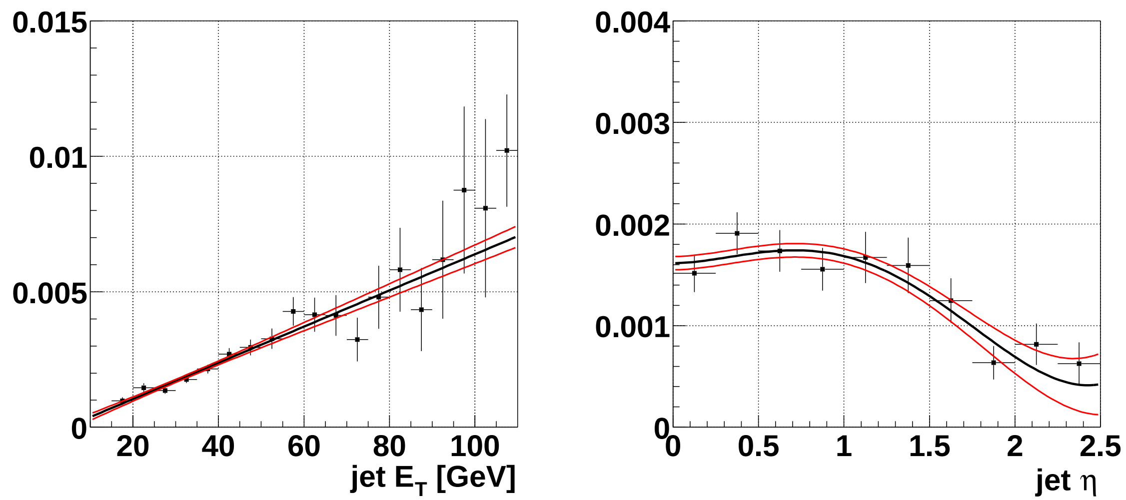


c-tagging efficiency in MC (CSIP)

Mis-tagging rate

- begin with negative inclusive tag rate ε^{-} measured in data
  - use negative side of DCA significance distribution (CSIP) or negative decay length (SVT)





mis-tagging rate (SVT)

- convert ε^{-} to light tag rate ε_l^{+} using scale factors determined on Monte Carlo: $SF_l = SF_{ll} \times SF_{hf} \sim 1$

  $SF_{ll} = \varepsilon_l^{+}/\varepsilon_l^{-}$ (long lived particles and fakes)

  $SF_{hf} = \varepsilon_l^{-}/\varepsilon^{-}$ (negative tag rate higher in HF jets)

Summary of tagging probabilities

CSIP		
	one tag	$\geq$2 tags
$t\bar{t} \to l+$jets	45.9$\pm$0.8%	15.8$\pm$0.3%
W+light jets	2.60$\pm$0.06%	0.03$\pm$0.01%

SVT		
	one tag	$\geq$2 tags
$t\bar{t} \to l+$jets	45.1$\pm$0.7%	13.9$\pm$0.1%
W+light jets	1.14$\pm$0.01%	$<$0.01%

probabilities to tag an e+jets event with at least 4 jets

W+jets background

- the number of W+jets events after tagging N_W^{tag} is related to the number of untagged W+jets events N_W as

$$N_W^{tag} = N_W P_W$$

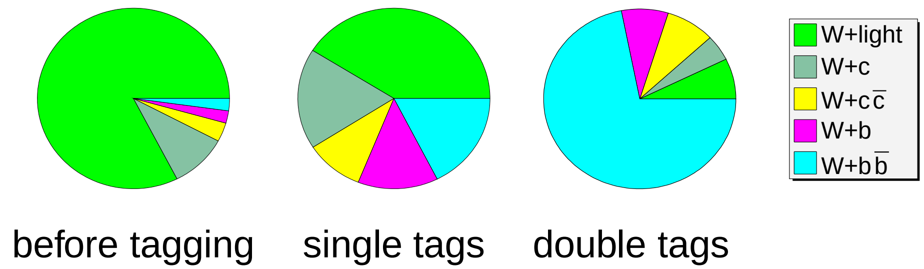
where P_W is average W+jets event tagging probability:

$$P_W = \sum_{flavor} F_{flavor} P_W(flavor)$$

- need to mix W+jets with different jet flavors in the right proportions

How to get F_{flavor} ?

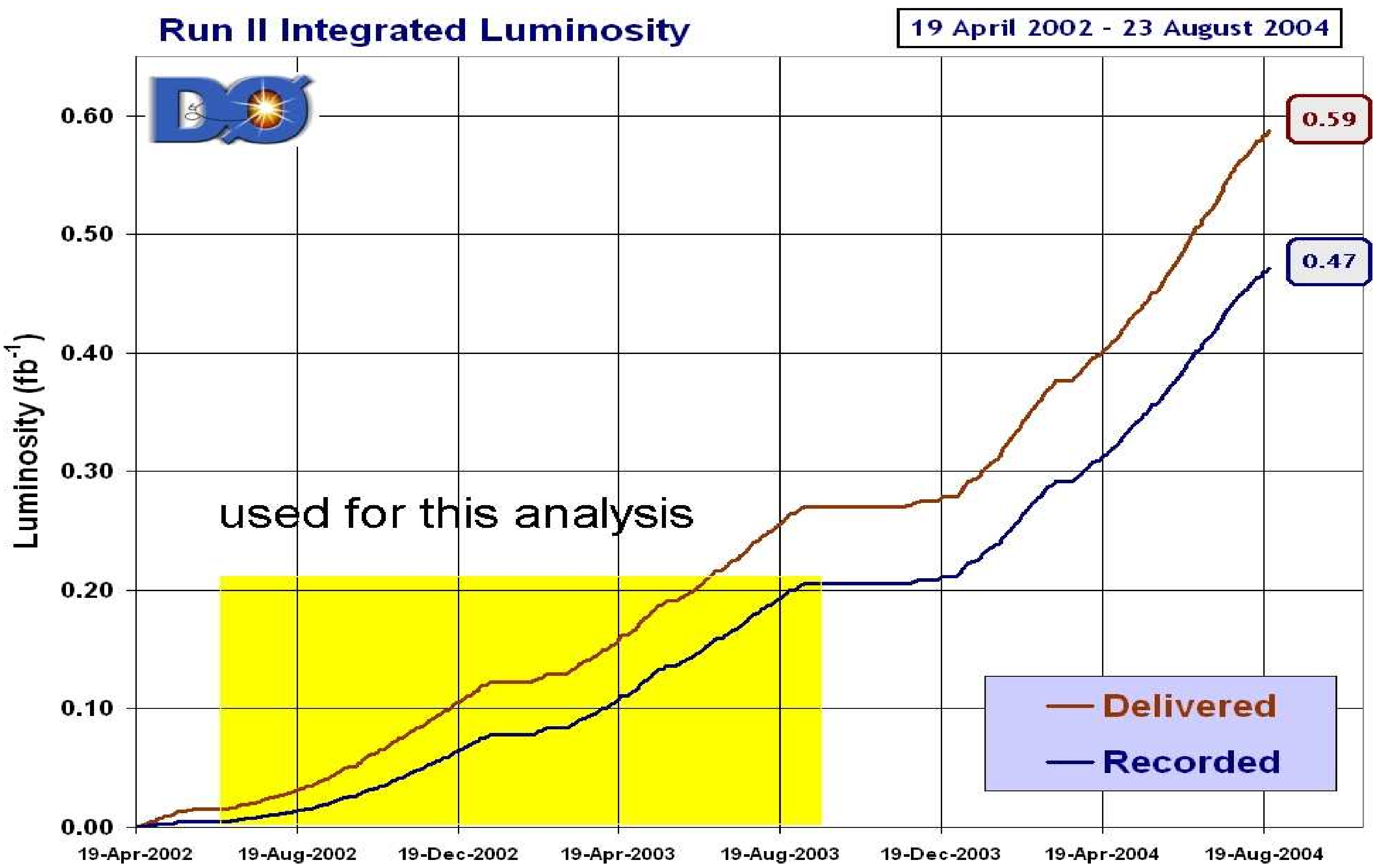
- use W+jets samples generated by ALPGEN 1.2 (CTEQ 6.1M) interfaced to PYTHIA 6.2 (CTEQ 5L)
- rely on the ratios of cross sections from Monte Carlo
- apply matching procedure to reduce double counting and sensitivity to parton generation cuts



Fractions of W+jets subprocesses with different flavors in e+$\geq$4 jets before and after tagging (CSIP)

Contribution	W+1jet	W+2jets	W+3jets	W+$\geq$4jets
$Wb\tilde{b}$		$(0.87 \pm 0.05)\%$	$(1.34 \pm 0.09)\%$	$(2.20 \pm 0.17)\%$
$Wc\tilde{c}$		$(1.11 \pm 0.08)\%$	$(2.11 \pm 0.16)\%$	$(3.43 \pm 0.46)\%$
$W(b\tilde{b})$	$(0.69 \pm 0.02)\%$	$(1.23 \pm 0.04)\%$	$(2.00 \pm 0.08)\%$	$(2.27 \pm 0.82)\%$
$W(c\tilde{c})$	$(1.10 \pm 0.05)\%$	$(1.91 \pm 0.07)\%$	$(2.71 \pm 0.18)\%$	$(3.4 \pm 1.2)\%$
Wc	$(4.50 \pm 0.17)\%$	$(6.80 \pm 0.27)\%$	$(7.21 \pm 0.36)\%$	$(5.30 \pm 0.35)\%$
$W + \text{jets(mistags)}$	$(93.7 \pm 2.8)\%$	$(88.1 \pm 2.8)\%$	$(84.6 \pm 3.4)\%$	$(83.4 \pm 2.5)\%$

Data sample: integrated luminosity



- this analysis: 169 pb^{-1} (e+jets), 158 pb^{-1} (μ+jets)
- now we have recorded $\sim$500 pb^{-1}

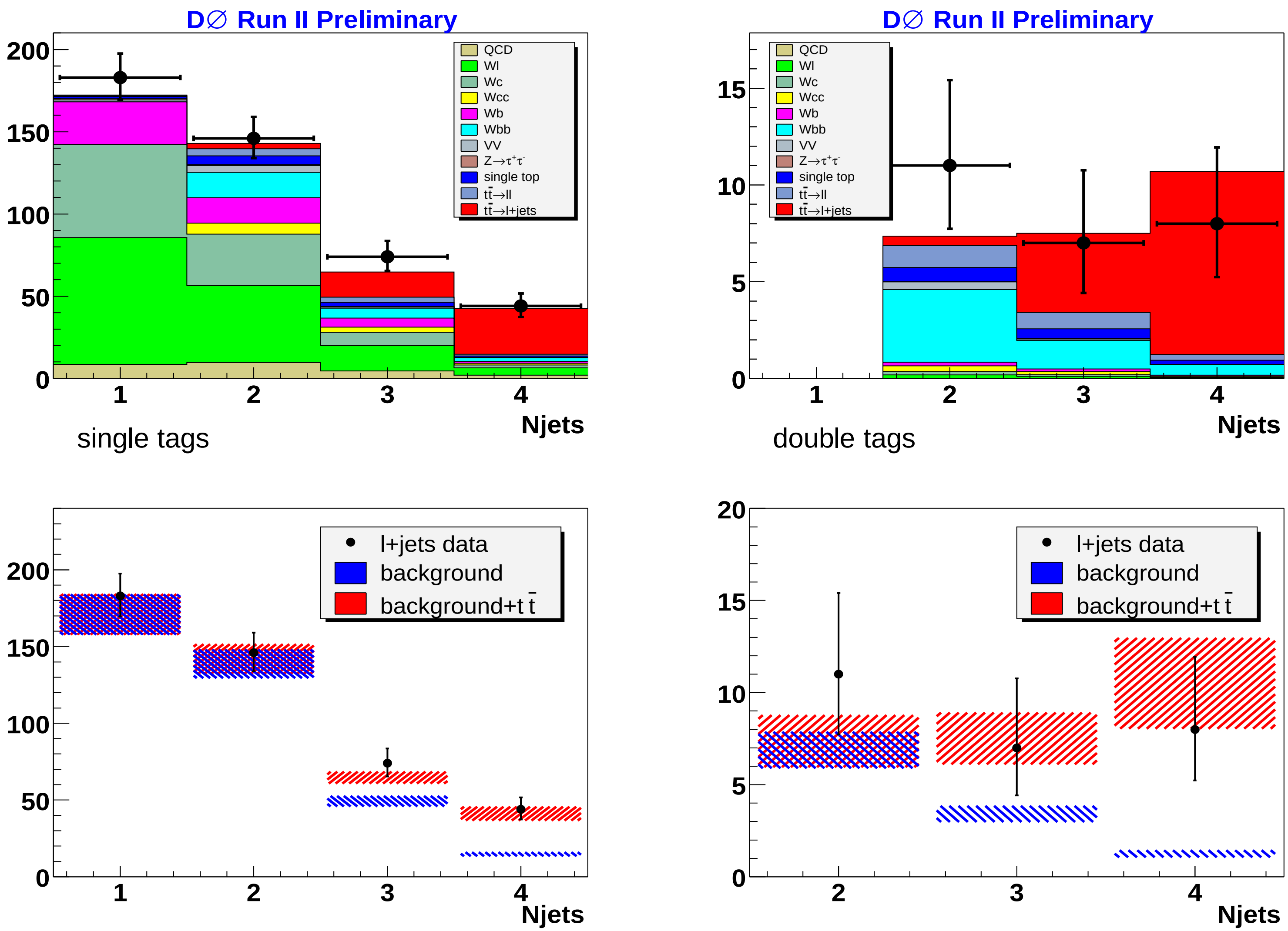
Data sample: the number of events

	W+1jet	W+2jets	W+3jets	W+$\geq$4jets
preselected	11586	4455	1105	295
CSIP: tagged	183	157	81	52
double tagged		11	7	8
SVT: tagged	119	128	76	49
double tagged		8	8	6

- work with four jet multiplicity bins
  - bins 1,2: use to control the background
  - bins 3,4: extract the $t\bar{t}$ production cross section
- cross section is extracted from a simultaneous likelihood fit to eight separate channels:
  - e+jets and μ+jets
  - events with 3 and $\geq$4 jets
  - single and double tagged events

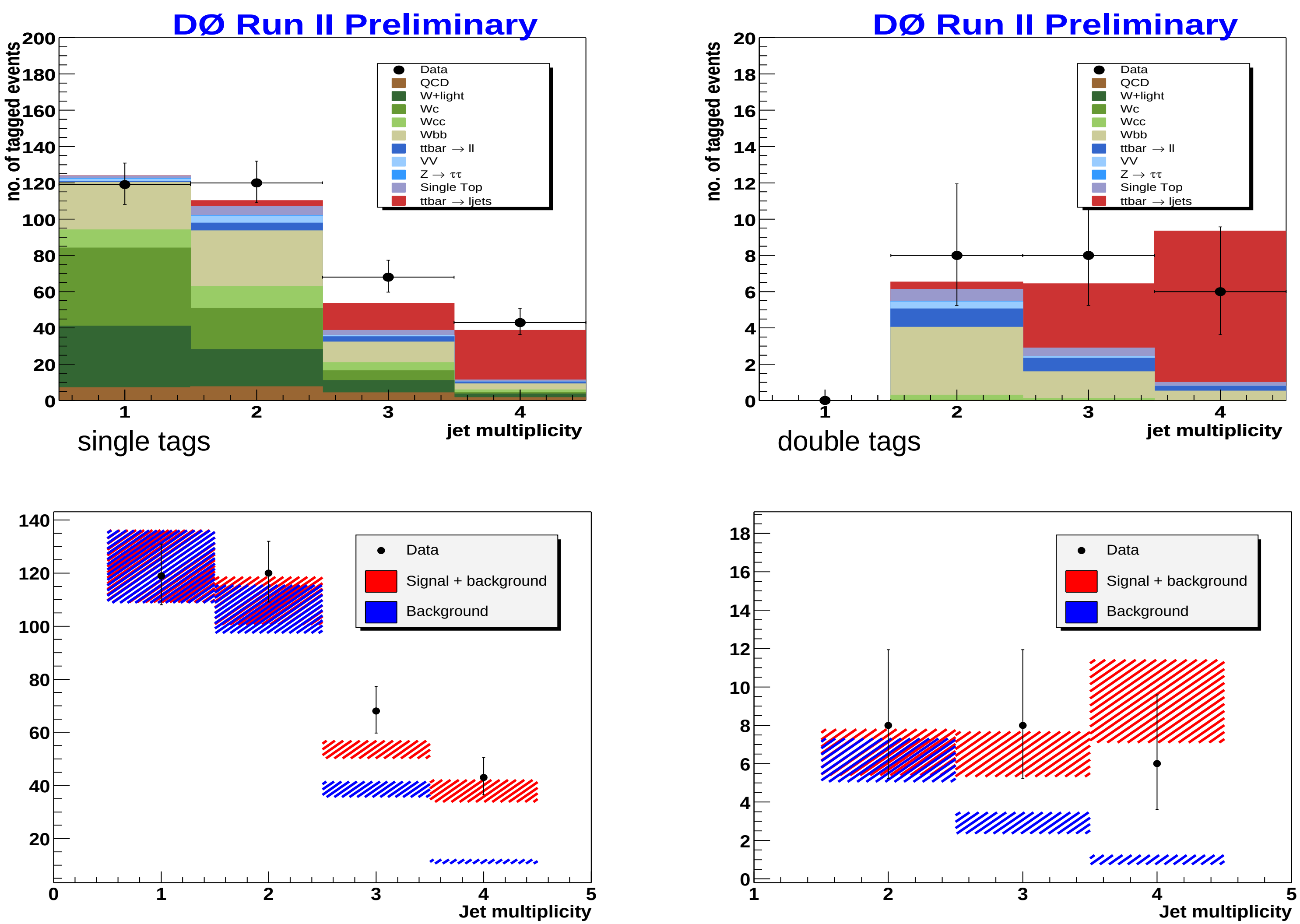
Observed vs predicted number of events: CSIP

signal prediction is shown for $\sigma_{t\bar{t}}$=7 pb



Observed vs predicted number of events: SVT

signal prediction is shown for $\sigma_{t\bar{t}}$=7 pb



The result

CSIP: $\sigma_{t\bar{t}} = 7.2^{+1.3}_{-1.2}\,(\mathrm{stat})\,^{+1.9}_{-1.4}\,(\mathrm{syst})\,\pm 0.5\,(\mathrm{lumi})\,\mathrm{pb}$

SVT: $\sigma_{t\bar{t}} = 8.2^{+1.3}_{-1.3}\,(\mathrm{stat})\,^{+1.9}_{-1.6}\,(\mathrm{syst})\,\pm 0.5\,(\mathrm{lumi})\,\mathrm{pb}$

theoretical prediction (NNLO): 6.77 ± 0.42 pb (hep-ph 0309045)

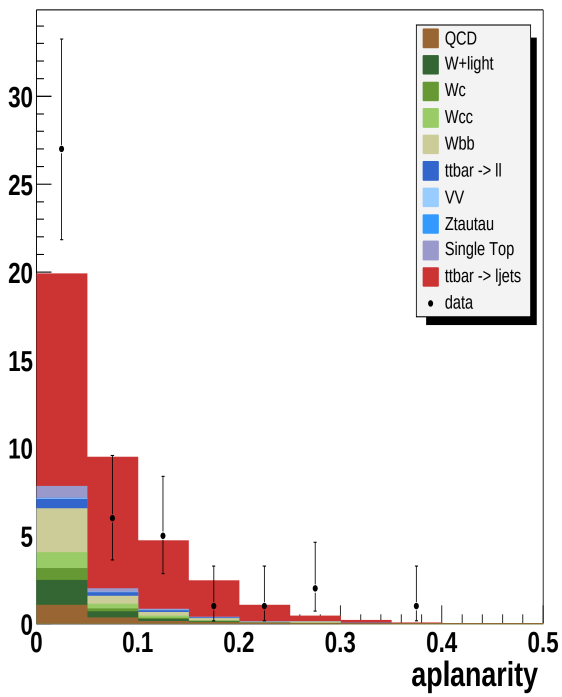
- correlations / combination of taggers under study
- for each tagger, 60% of tags are also found by another one

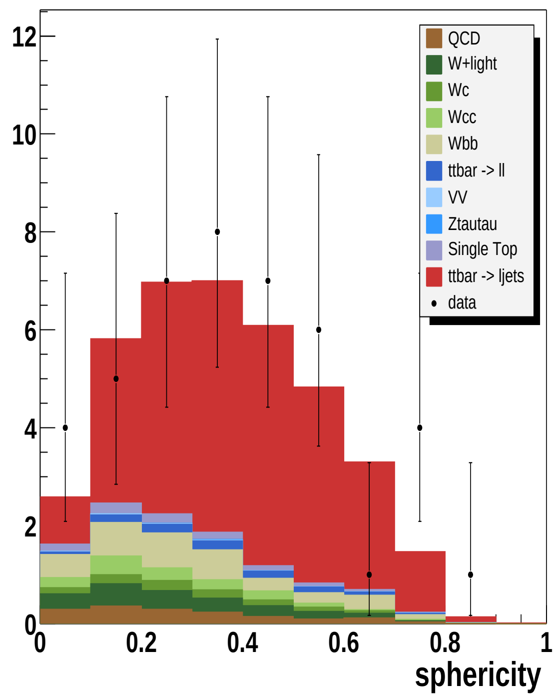
Sources of systematics

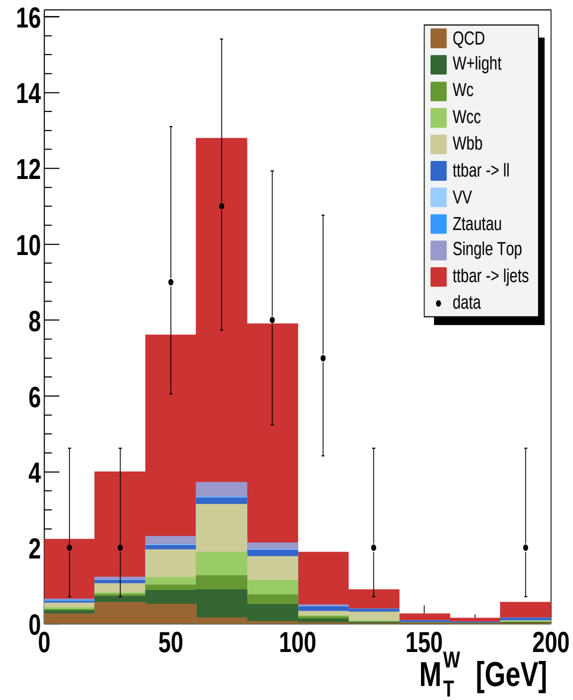
- main sources of systematics:
  - jet energy scale, $\Delta\sigma_{t\bar{t}} \sim 1$ pb
  - b-tagging efficiency in data, $\Delta\sigma_{t\bar{t}} \sim 0.9$ pb
  - W fractions, $\Delta\sigma_{t\bar{t}} \sim 0.8$ pb
- full list of systematic uncertainties:

- W fractions
- trigger efficiency
- primary vertex efficiency
- matrix method efficiencies
- object ID efficiency
- jet resolution
- jet energy scale
- heavy flavor tagging efficiency in MC
- semileptonic b-tagging efficiency in data
- taggability
- negative tag rate and light flavor SF
- fragmentation model
- assumption $SF_b = SF_c$
- top mass

Cross-checks: aplanarity, sphericity, transverse W mass

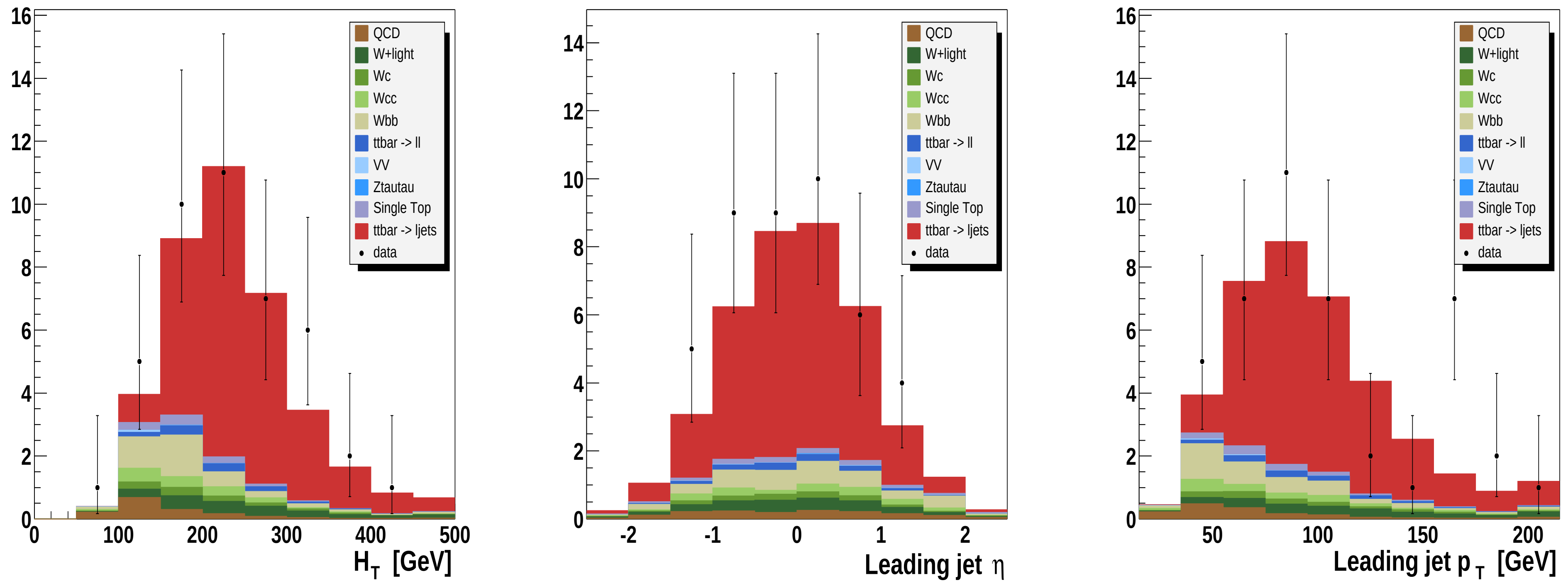






shown for SVT

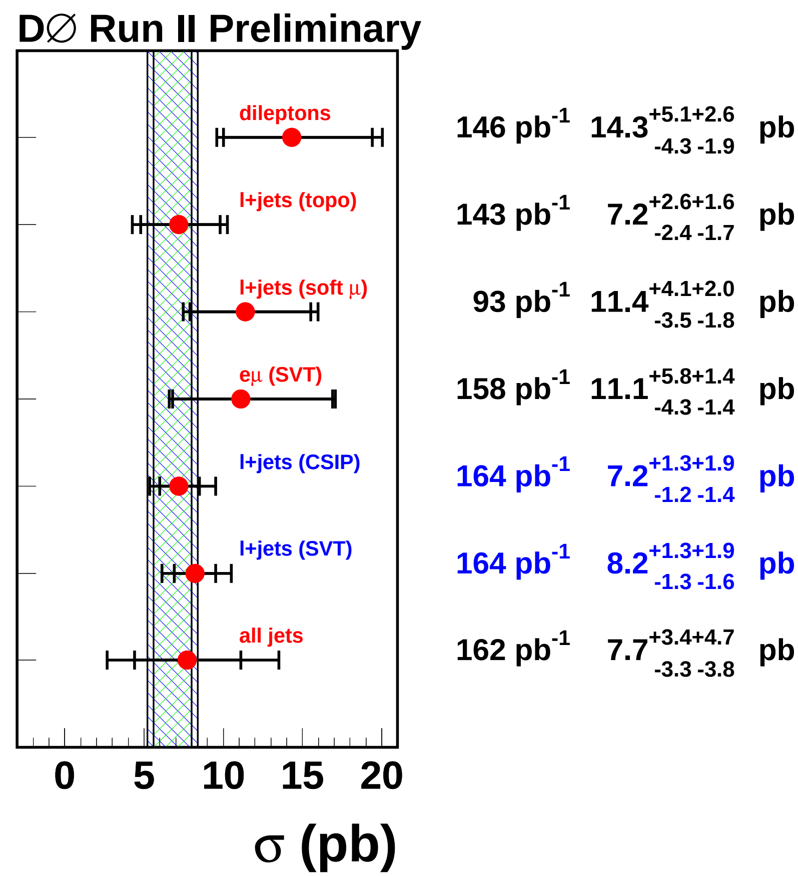
Cross-checks: scalar sum of jet energies, leading jet η and p_T



shown for SVT

Conclusions

- we have measured the $t\bar{t}$ production cross section in lepton+jets channel with lifetime b-tagging
- two different methods were used to cross-check results
- this is the most precise measurement of the $t\bar{t}$ production cross section in DØ
- the result is in a good agreement with the Standard Model prediction



- the quoted systematic error is conservative, expect significant improvement
- now have 2-3 times more data, can reduce statistical and many systematic errors